DOCUMENT RESUME

ED 038 897 FL 001 647

AUTHOR	Savard, Jean-Guy; Richards, Jack C.
TITLE	Measuring the Utility of a Basic French Vocabulary.
INSTITUTION	International Centre for Research on Bilingualism, Quebec (Canada).
PUB DATE	69
NOTE	13p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$0.75
DESCRIPTORS	Classification, *Computational Linguistics, Dictionaries, *French, Indexes (Locaters), Indexing, Language, Measurement Techniques, Subject Index Terms, Technical Reports, *Vocabulary, *Word Frequency, *Word Lists, Written Language

ABSTRACT

Using frequency, range, availability, and coverage as objectively weighted indices of utility in the development of a basic French vocabulary list, the authors describe the technical procedure used in this study and list the resultant first 153 words according to the four indices and the utility rating. Weighting factors for each index, based on extant studies ("Le Francais Fondamental", C.R.E.D.I.F. [Centre de Recherche et d'Etude pour la Diffusion du Francais] studies, others by Mackey and Savard), are clarified in detail. A bibliography is included. (RL)

EDO 38897

centre international de recherches sur le bilinguisme

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

MEASURING THE UTILITY OF A BASIC FRENCH VOCABULARY

by JEAN-GUY SAVARD

and JACK C. RICHARDS

1969

FL 001 947

CIRB
ICRB

Documentation internationale et recherche interdisciplinaire dans le domaine du bilinguisme, du biculturalisme et des phénomènes connexes

International documentation and interdisciplinary research in the field of bilingualism, biculturalism and related phenomena

---

Centre international de recherches sur le bilinguisme
International Centre for Research on Bilingualism
Université Laval, Cité universitaire
Ste-Foy 10e (Québec) Canada

ED038897

MEASURING THE UTILITY OF A BASIC FRENCH VOCABULARY

by Jean-Guy Savard
and Jack C. Richards

International Center for Research on Bilingualism

To appear in *Te Reo* 1969

MEASURING THE UTILITY OF A BASIC FRENCH VOCABULARY

Jean-Guy Savard and Jack C. Richards

International Center for Research on Bilingualism,

Laval University, Quebec, Canada

The description of language and the teaching of languages are separate, but related activities. Whereas the linguist is concerned chiefly with his subject matter, the language teacher is concerned with the relationship between himself, the subject matter and the learner. Theories of language teaching differ according to the emphasis they give to each of these variables.

Faced with the question of what should be taught, teachers of foreign languages have been advised to make use of word frequency counts. For the linguist, the most frequently occurring words in a language may be thought of as a structural statement representative of the language as a whole[1]. A frequency list may be designed to act as a compressed body of evidence in which all the features of the parent corpus are preserved. For the teacher, frequency counts offer a practical plan for language instruction, since they enable the total lexical corpus of the language, perhaps 500,000 words, to be reduced to a managable size that can operate effectively in terms of some limited objective. This might be a 4,000 word reading vocabulary, a 2,000 word speaking vocabulary, or a vocabulary for some other specified task. The notion of word counting is at least as old as classical times. "...all the twentieth century added to known principles were accurate mathematical procedures, the computer,

and a revised set of technical terms"[2].

In recent years, additional criteria for vocabulary selection have been introduced or old notions refined, as the acceleration of modern foreign language teaching, created the demand (in theory at least) for objective measures by which the selection of materials for language instruction could proceed. The concept of range, quantified for a number of the world's principal languages, indicates different degrees of power for words, based on their distribution thoughout a corpus rather than on their raw frequency relative to the rest of the words in a list.

Frequency and range however, still emphasize the linguistic rather than the pedagogic structure of vocabulary, if one thinks of the latter in terms of the words needed to teach a language and the former as the words which are most often needed by the language in non-didactic circumstances. That the two require a different analysis is evident when we examine a frequency count in detail.

To teach a language, it is usual to recreate the contexts in which words are used, choosing situations which the learner is likely to encounter. Yet the vocabulary of many familiar situations does not rank highly in frequency lists. Soap, bath, cushion, chalk, stomach are not within the first 2,000 words of Thorndike and Lorge's list[3]. In the general frequency list of français fondamental[4] blouse, brique and tablier have frequencies of 2, 1, and 3 respectively. Many concrete nouns of high teachability are given low priority in frequency lists, and this does not change significantly according to the type of material counted.

The reasons for this are found within the nature of a

frequency list itself. Michéa notes that the effect of a frequency count is to reduce a corpus to a set of frequencies in which the value of a given frequency is necessarily relative to that of the other frequencies[5]. Since the words at the top of a list occur with a very high frequency, (the first 250 words representing as much as 80% of a text) there is only a small percentage left to be shared by the thousands of other words in the language. The most frequent words in the language are words with grammatical, abstract or general meaning. The grammatical words are frequent because it is impossible to produce a sentence without them. Both grammatical and abstract words have a variety of meanings and hence several opportunities to occur in any given text. Infrequent words are generally words with fewer meanings. They are highly context-bound, carry more information, and are likely to vary from one topic of discourse to another.

To measure the importance of concrete nouns, the concept of availability (disponibilité) was developed by the elaborators of fundamental French, in an attempt to rate nouns according to their degree of association within specific categories and situations. Using categories such as "food", "the house", "the school", students were asked to write down the words which first came to mind when they thought of these categories. The lists of nouns so obtained were used to correct the frequency lists used for français fondamental, the French government-supported project for the teaching of French as a second language. An alternative method for measuring the significance of concrete nouns has been developed by Richards[6] using students' estimations of their usage of words. Figures obtained reflect the general cultural importance of nouns, and are likewise being used in conjunction with frequency lists for designing vocabulary lists for second language teaching[7].

Teachers have acknowledged the need for additional criteria for vocabulary selection however, and lists such as A General Service List of English Words[8] utilize intuitively the notion of coverage (valence) to justify the inclusion or exclusion of items from word lists. Coverage is an attempt not to find out what words do, but to measure the work words can do. The usefulness of some members of the lexicon in Basic English[9] for example, is determined not by frequency of occurrence or availability, but by the degree of extension of meaning which some words have, and the degree to which words can replace others through definition. Seat includes the meaning of chair, bench, stool and can be used instead of these words. Make may help define scare (makes me frightened) retard (makes me slower) and so on, and so can help replace words such as these.

The existence of these four criteria for vocabulary selection, frequency, range, availability and coverage, prompted the study described here, which was an attempt to arrive at a basic French word list using objectively weighted multiple indices.[10] Our aim was to compare the ratings given to a basic French vocabulary by these four indices and to merge them into a single factor which we have called "utility". In the word list which emerges, utility means simply that a word is useful according to the degree to which it rates highly on each of these factors.

Sources for the study.

The indices for frequency and range utilized in the study are those established for français fondamental. The index of availability is a combined index, utilizing data from the C.R.E.D.I.F. study which produced français fondamental and from a study by Mackey and Savard conducted with French speakers in

New Brunswick, Canada, between 1963 and 1967[11]. The index of coverage is taken from Savard[12], and is itself a combined index based on (a) a measure of the degree of defining power of the vocabulary according to the number of times each word was used in dictionary definition; (b) a measure of the degree to which the vocabulary could replace other words through synonymity, based on thesaurus analysis; (c) a measure of extension, or the power of a word to multiply its meanings, arrived at through a count of the number of meanings listed for each word in Robert[13], and (d) a measure of combination power, that is, the power of words to form new lexical items by combining with other words, again based on counts of word combinations in dictionaries[14].

This gave a total of around 3,300 words, excluding structural words, for which comparative ratings were available. We then normalized the four indices to make them mutually comparable. Normalization was achieved by applying the formula

$$Z = \frac{X - M}{\sigma}$$

where X = the rating for a word, M = the mean, and σ = the standard deviation. Transforming the four ratings by this formula gave for each word, a scale varying between 0 and 800, with a theoretical mean of 400. The maximum scale on the combined rating so formed - the utility rating - is 3,200. Subsequently the coefficient of correlation between each of the indices was determined, and the results factor analyzed to determine the extent to which each constituent of utility measured something different. This showed that fundamentally different properties of a word were measured under frequency, availability and coverage, while range and frequency were more closely related for the corpus studied[15].

The results of the first 153 words of the utility list for basic French are shown in table I.

TABLE I. TOP 153 WORDS AS MEASURED BY UTILITY

RANK	WORD	RANGE	FREQUENCY	AVAILABILITY	COVERAGE	UTILITY
1	FAIRE	628.7	628.7	628.7	601.3	2487.6
2	MAISON	603.7	603.7	574.1	592.9	2374.3
3	PARLER	628.7	628.7	484.6	581.7	2323.7
4	PASSER	628.7	628.7	462.1	610.9	2330.4
5	VOULOIR	628.7	628.7	477.8	590.3	2325.6
6	ALLER	628.7	628.7	462.1	601.3	2321.0
7	PRENDRE	628.7	628.7	443.7	619.4	2320.6
8	METTRE	628.7	628.7	462.1	599.0	2318.7
9	PETIT	628.7	628.7	462.1	595.4	2315.0
10	TETE	559.7	559.7	594.2	598.9	2312.5
11	PIED	568.9	559.7	628.7	551.8	2309.1
12	MAIN	568.9	568.9	628.7	541.5	2308.1
13	VENIR	628.7	628.7	443.7	587.4	2288.6
14	POUVOIR	628.7	628.7	462.1	567.0	2286.7
15	JOUR	628.7	628.7	484.6	544.3	2286.4
16	TRAIN	548.1	564.1	628.7	545.4	2286.3
17	DIRE	628.7	628.7	443.7	584.1	2285.3
18	SAS	555.6	541.3	628.7	559.6	2285.3
19	GRAND	628.7	616.6	443.7	593.3	2282.3
20	SAVOIR	628.7	628.7	443.7	580.7	2281.8
21	VOIR	628.7	628.7	443.7	577.8	2278.9
22	DONNER	628.7	616.6	443.7	584.2	2273.2
23	TROUVER	628.7	628.7	443.7	570.7	2271.9
24	ECOLE	579.9	586.5	574.1	530.4	2270.9
25	DEVOIR	628.7	628.7	471.3	537.3	2266.2
26	CHEVAL	532.2	574.1	628.7	531.0	2266.1
27	TENIR	628.7	603.7	443.7	581.8	2257.8
28	TRAVAILLER	579.9	594.2	507.7	572.8	2254.6
29	EAU	564.1	579.9	564.1	541.9	2250.0
30	BON	628.7	603.7	443.7	572.3	2248.4
31	CROIRE	628.7	628.7	462.1	526.0	2245.6
32	CHOSE	628.7	628.7	462.1	524.8	2244.5
33	HOMME	574.1	574.1	481.2	612.1	2241.5
34	PARTIR	603.7	603.7	443.7	589.8	2240.8
35	ENFANT	594.2	603.7	488.1	552.6	2238.6
36	TEMPS	628.7	616.6	443.7	545.3	2234.3
37	BOIS	544.6	574.1	594.2	518.8	2231.8
38	VOITURE	559.7	586.5	559.7	521.4	2227.3
39	TRAVAIL	586.5	594.2	481.2	565.1	2227.0
40	ARRIVER	628.7	628.7	443.7	525.4	2226.6
41	TABLE	532.2	538.2	628.7	527.3	2226.4
42	MANGER	579.9	594.2	477.8	573.6	2225.5
43	BEAU	594.2	594.2	462.1	573.0	2223.5
44	CONNAITRE	628.7	603.7	443.7	545.6	2221.7
45	MARCHER	586.5	586.5	471.3	574.5	2218.9
46	FEMME	586.5	603.7	503.5	522.3	2215.9
47	PIECE	544.6	564.1	512.1	593.5	2214.3
48	LIT	535.1	544.6	628.7	503.0	2211.5
49	DEMANDER	628.7	603.7	443.7	533.8	2209.9
50	SORTIR	603.7	594.2	443.7	567.7	2209.2
51	TERRE	544.6	551.7	526.7	585.6	2208.7

52	RUE	603.7	574.1	538.2	492.0	2207.9
53	FOIS	628.7	616.6	443.7	515.4	2204.4
54	PLACE	574.1	574.1	521.6	534.4	2204.2
55	COEUR	529.4	532.2	628.7	509.3	2199.7
56	ENTENDRE	594.2	586.5	443.7	575.2	2199.6
57	FALLOIR	628.7	628.7	443.7	497.4	2198.6
58	REVENIR	594.2	586.5	443.7	571.2	2195.6
59	HEURE	628.7	628.7	443.7	494.2	2195.3
60	PERSONNE	579.9	574.1	495.5	545.0	2194.6
61	VRAI	594.2	594.2	443.7	559.5	2191.6
62	GROS	586.5	586.5	443.7	573.6	2190.3
63	CHAMBRE	555.6	564.1	594.2	476.3	2190.2
64	PERDRE	586.5	579.9	443.7	575.0	2185.1
65	MONDE	568.9	559.7	495.5	559.1	2183.2
66	ECRIRE	564.1	559.7	497.5	559.4	2180.7
67	REGARDER	586.5	586.5	443.7	562.2	2178.9
68	COUP	586.5	579.9	443.7	568.4	2178.5
69	HISTOIRE	579.9	586.5	497.5	513.9	2177.8
70	PORTE	541.3	532.2	579.9	521.1	2174.5
71	BUREAU	529.4	532.2	628.7	483.0	2173.4
72	AIMER	603.7	603.7	443.7	522.2	2173.2
73	MOMENT	628.7	603.7	443.7	496.8	2172.9
74	HAUT	555.6	535.1	514.4	567.5	2172.6
75	COMPRENDRE	603.7	603.7	443.7	520.0	2171.0
76	LIVRE	491.8	538.2	628.7	512.1	2170.8
77	MONTER	579.9	579.9	443.7	566.7	2170.2
78	PENSER	603.7	594.2	443.7	526.3	2167.8
79	VIEUX	586.5	594.2	443.7	542.0	2166.4
80	FILLE	574.1	586.5	462.1	543.5	2166.3
81	ARRETER	579.9	568.9	462.1	550.8	2161.8
82	RENDRE	586.5	586.5	443.7	543.3	2160.0
83	CHIEN	507.7	535.1	628.7	488.3	2159.9
84	TOMBER	574.1	564.1	443.7	577.0	2158.9
85	JOUER	559.7	579.9	477.8	541.1	2158.6
86	DAME	555.6	544.6	555.6	500.8	2156.5
87	FORT	579.9	559.7	462.1	553.9	2155.7
88	OEIL, YEUX	535.1	544.6	524.1	547.0	2150.8
89	CAFE	535.1	551.7	564.1	499.9	2150.8
90	SERVICE	568.9	568.9	462.1	550.3	2150.3
91	RESTER	603.7	594.2	443.7	507.5	2149.1
92	BATEAU	488.1	519.1	628.7	512.2	2148.2
93	MAUVAIS	568.9	555.6	443.7	580.0	2148.2
94	BRAS	524.1	512.1	628.7	483.2	2148.1
95	PARTIE	559.7	548.1	443.7	593.3	2144.8
96	VIE	586.5	579.9	443.7	532.3	2142.4
97	MUR	512.1	519.1	628.7	481.1	2141.0
98	COMMENCER	594.2	586.5	443.7	516.6	2141.0
99	MAITRE	516.7	497.5	555.6	570.0	2139.7
100	GENS	594.2	603.7	471.3	469.8	2139.0
101	EXEMPLE	603.7	594.2	443.7	497.2	2138.8
102	NEZ	507.7	493.6	628.7	508.5	2138.5

103	PAYER	568.9	579.9	443.7	545.8	2138.3
104	OCCUPER	568.9	564.1	443.7	560.5	2137.2
105	SEUL	594.2	586.5	443.7	512.2	2136.6
106	LAISSER	594.2	586.5	443.7	512.1	2136.5
107	VENDRE	535.1	564.1	501.4	534.8	2135.5
108	FACON	574.1	574.1	443.7	542.9	2134.9
109	PERE	564.1	574.1	471.3	524.5	2134.1
110	IDEE	574.1	564.1	443.7	551.8	2133.8
111	FINIR	586.5	574.1	443.7	529.4	2133.7
112	POINT	579.9	579.9	443.7	529.0	2132.6
113	THEATRE	516.7	529.4	541.3	542.1	2129.5
114	SEULEMENT	603.7	586.5	443.7	495.4	2129.2
115	AN	628.7	616.6	443.7	439.5	2128.5
116	AMI	568.9	568.9	471.3	519.3	2128.4
117	COUCHER	548.1	551.7	484.6	543.1	2127.5
118	APPELER	594.2	594.2	443.7	495.2	2127.3
119	PAPIER	529.4	541.3	555.6	498.6	2124.9
120	JAMBE	497.5	505.6	628.7	493.0	2124.8
121	ATTENDRE	586.5	579.9	443.7	514.6	2124.7
122	PORTER	551.7	548.1	443.7	581.0	2124.5
123	CHERCHER	594.2	579.9	462.1	487.7	2124.0
124	FIN	559.7	551.7	462.1	548.8	2122.4
125	LIRE	555.6	579.9	507.7	476.1	2119.3
126	CUISINE	535.1	535.1	594.2	454.6	2119.0
127	ARGENT	551.7	564.1	462.1	541.0	2119.0
128	MOT	555.6	579.9	443.7	539.7	2118.9
129	CHANGER	579.9	564.1	443.7	530.7	2118.4
130	TOUCHER	541.3	538.2	471.3	566.9	2117.7
131	TOURNER	544.6	551.7	443.7	577.2	2117.2
132	ROUTE	541.3	568.9	514.4	491.4	2115.9
133	TIRER	538.2	535.1	443.7	598.9	2115.8
134	TOUR	529.4	541.3	462.1	582.7	2115.6
135	MOURIR	524.1	541.3	443.7	606.3	2115.4
136	FRANC	586.5	603.7	443.7	481.0	2114.8
137	SENTIR	559.7	564.1	443.7	546.9	2114.4
138	PARAITRE	579.9	579.9	443.7	510.8	2114.3
139	LANGUE	512.1	512.1	568.9	520.8	2113.9
140	PLAT	516.7	507.7	586.5	501.4	2112.2
141	COUPER	535.1	541.3	481.2	553.8	2111.4
142	GENERAL	574.1	568.9	462.1	506.1	2111.2
143	LEVER	535.1	532.2	495.5	547.2	2110.1
144	SERVIR	568.9	564.1	443.7	532.4	2109.1
145	BOUT	574.1	568.9	443.7	521.5	2108.2
146	LONG	574.1	564.1	443.7	525.7	2107.6
147	VIVRE	559.7	555.6	443.7	548.3	2107.2
148	ECOUTER	586.5	586.5	462.1	471.3	2106.4
149	DERNIER	594.2	586.5	443.7	481.9	2106.3
150	VILLE	559.7	574.1	477.8	494.6	2106.3
151	JUSTE	579.9	555.6	443.7	526.8	2106.0
152	RAPPELER	574.1	574.1	443.7	513.2	2105.2
153	TABLEAU	465.2	481.2	628.7	529.3	2104.4

FOOTNOTES

I. Juilland and Rodriguez, I964: XXIII.

2. Kelly, I969: I84.

3. Thorndike and Lorge, I944.

4. Gougenheim et al, I964.

5. Michéa, I964 : I9-33.

6. Richards, J, I969 a,b.

7. Richards, J, I969 c.

8. West, I953.

9. Richards,I.A,I943.

I0. Savard and Richards. I970.

II. Mackey and Savard, (in press)

I2. Savard, (in press)

I3. Robert, I965.

I4. Mackey and Savard, I967.

I5. See Savard and Richards, chapter 4-5.

BIBLIOGRAPHY

Gougenheim,G.et al, I964. L'elaboration du français fondamental. Paris, Didier.

Juilland, A and Rodriguez, C.E, I964. Frequency Dictionary of Spanish Words.The Hague, Mouton.

Kelly,L.G, I969. 25 Centuries of Language Teaching. Massachusetts, Newbury House.

Mackey,W.F, and Savard,Jean-Guy, I967. "The Indices of Coverage", International Review of Applied Linguistics, V/2-3, 7I-I2I.

Mackey,W.F, and Savard,Jean-Guy et al, (in press). Le vocabulaire disponible en France et en Acadie.

Michéa,R, I964. "Basic Vocabularies", New Research and Techniques for the Benefit of Modern Language Teaching. Strasbourg, Council for Cultural Cooperation. I9-33.

Richards,Jack,I969 a. "Word Association, Word Availability, and Word Familiarity as Selection Factors in an English Base Vocabulary". Paper read at the annual meeting of the Canadian Linguistic Association, York University, Toronto, June I969.

Richards, Jack, I969 b. "A Psycholinguistic Measure of Vocabulary Selection". Research Paper, International Center for Research on Bilingualism, Laval University ,Quebec, August I969.

Richards, Jack,I969 c. "A Familiarity List of English Nouns". Research Monograph, International Center for Research on Bilingualism, December I969.

Richards, I.A, I943. Basic English and its Uses. London, Kegan Paul.

Robert,P, I965. Dictionaire alphabétique et analogique de la langue française. Paris, Nouveau Littré (6 vols).

Savard, Jean-Guy (in press) La Valence Lexicale.

Savard ,Jean-Guy and Richards, Jack, I970.Les indices d'utilité du vocabulaire français fondamental.Quebec, Les Presses de l'Université Laval.

Thorndike,E.W, and Lorge,I, I944. The Teachers Word Book of 30,000 Words.New York, Columbia Teacher's College.

West,M.I953. A General Service List of English Words. London, Longmans.